

Metadata Tools

Ted Habermann, NOAA/NESDIS/NGDC
NCDC Metadata Workshop

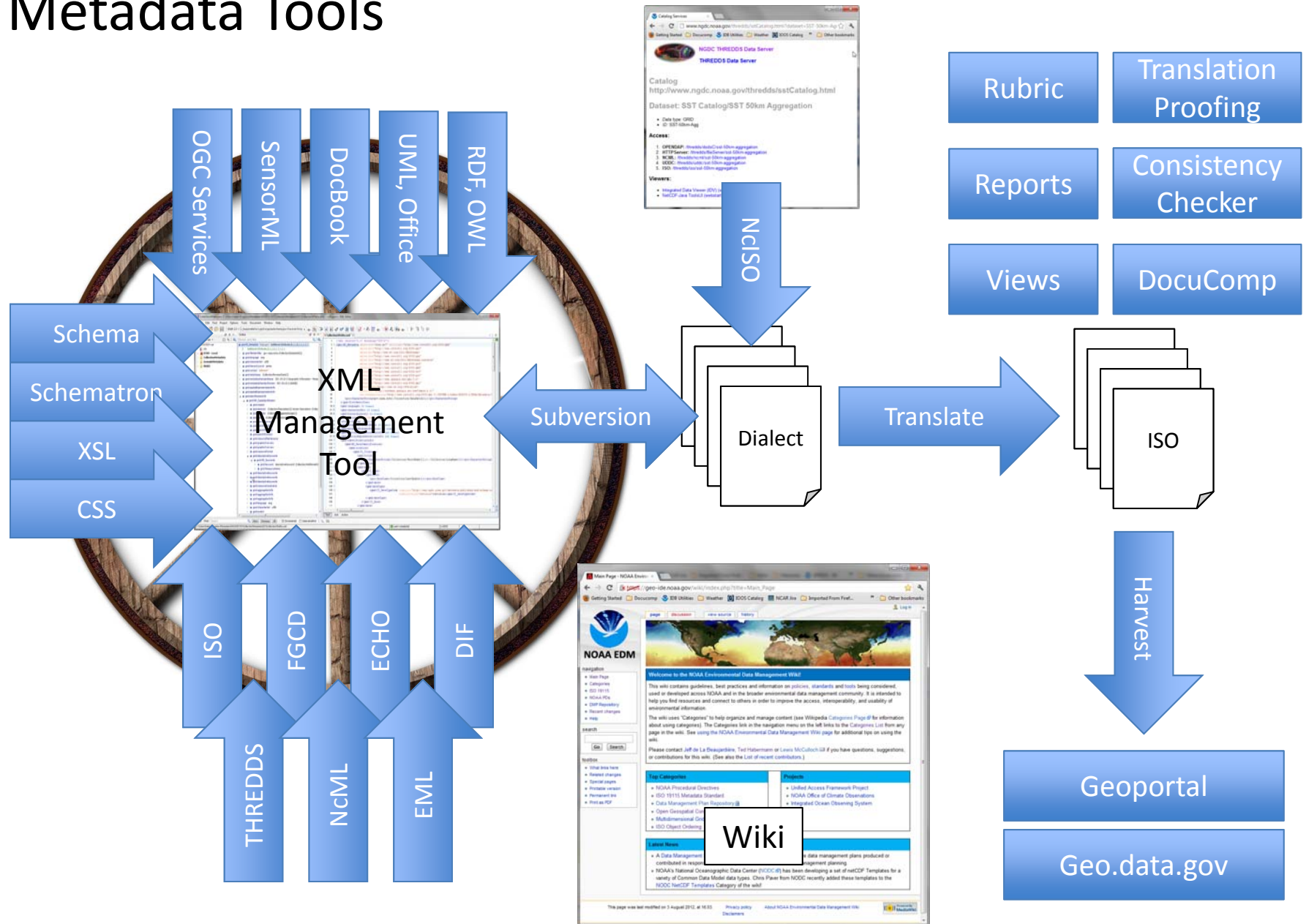


The reason why innovation often seems to be so difficult for **established** firms is that they employ highly capable people and then set them to work within processes and values that weren't designed to facilitate success with the task at hand.

Ensuring that capable people are ensconced in capable organizations is a major management responsibility in an age such as ours, when the ability to cope with accelerating change has become so critical.

Clayton Cristensen, Coping With Your Organization's Innovation Capabilities, in Leading For Innovation and Organizing for Results

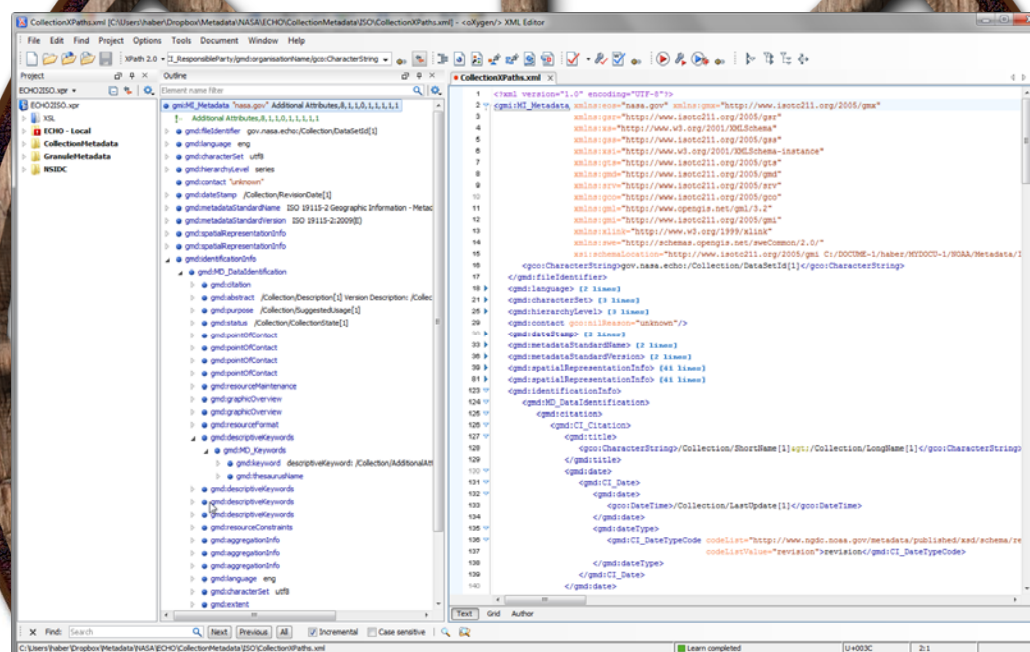
Metadata Tools



XML Editor – The Tool at the Hub

Schema

Schematron



Subversion

XSL

XML Management Tool

The screenshot displays the XML Management Tool interface, which is divided into several panes. The top pane shows the project structure with folders like 'ECHO2ISO.xpr', 'XSL', 'ECHO - Local', 'CollectionMetadata', 'GranuleMetadata', and 'NSIDC'. The middle pane shows the 'Outline' view with a tree of elements including 'gmi:MI_Metadata', 'gmd:fileIdentifier', 'gmd:language', 'gmd:characterSet', 'gmd:hierarchyLevel', 'gmd:contact', 'gmd:resourceMaintenance', 'gmd:graphicOverview', 'gmd:resourceFormat', 'gmd:descriptiveKeywords', 'gmd:MD_Keywords', 'gmd:keyword', 'gmd:extent', and 'gmd:CI_DateTypeCode'. The right pane shows the XML code editor with the following code:

```
<?xml version="1.0" encoding="UTF-8"?>
<gmi:MI_Metadata xmlns:eos="nasa.gov" xmlns:gmw="http://www.isotc211.org/2005/gmw"
  xmlns:gtr="http://www.isotc211.org/2005/gtr"
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  xmlns:gss="http://www.isotc211.org/2005/gss"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:gts="http://www.isotc211.org/2005/gts"
  xmlns:gmd="http://www.isotc211.org/2005/gmd"
  xmlns:srv="http://www.isotc211.org/2005/srv"
  xmlns:gco="http://www.isotc211.org/2005/gco"
  xmlns:gml="http://www.opengis.net/gml/3.2"
  xmlns:gmi="http://www.isotc211.org/2005/gmi"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:swe="http://schemas.opengis.net/sweCommon/2.0/"
  xsi:schemaLocation="http://www.isotc211.org/2005/gmi C:/DOCUME~1/haer/MYDOCU~1/NOAA/Metadata/1
  <gco:CharacterString>gov.nasa.echo:/Collection/DataSetId[1]</gco:CharacterString>
  </gmd:fileIdentifier>
  <gmd:language> [2 lines]
  <gmd:characterSet> [3 lines]
  <gmd:hierarchyLevel> [3 lines]
  <gmd:contact gco:nilReason="unknown"/>
  <gmd:dateStamp> [2 lines]
  <gmd:metadataStandardName> [2 lines]
  <gmd:metadataStandardVersion> [2 lines]
  <gmd:spatialRepresentationInfo> [41 lines]
  <gmd:spatialRepresentationInfo> [41 lines]
  <gmd:identificationInfo>
    <gmd:MD_DataIdentification>
      <gmd:citation>
        <gmd:title>
          <gco:CharacterString>/Collection/DataSetId[1]</gco:CharacterString>
        </gmd:title>
        <gmd:date>
          <gmd:CI_Date>
            <gmd:date>
              <gco:DateTime>/Collection/DataSetId[1]</gco:DateTime>
            </gmd:date>
            <gmd:dateType>
              <gmd:CI_DateTypeCode>
                <gco:CodeListValue>revision</gco:CodeListValue>
              </gmd:CI_DateTypeCode>
            </gmd:dateType>
          </gmd:CI_Date>
        </gmd:date>
      </gmd:citation>
    </gmd:MD_DataIdentification>
  </gmd:identificationInfo>
</gmi:MI_Metadata>
```

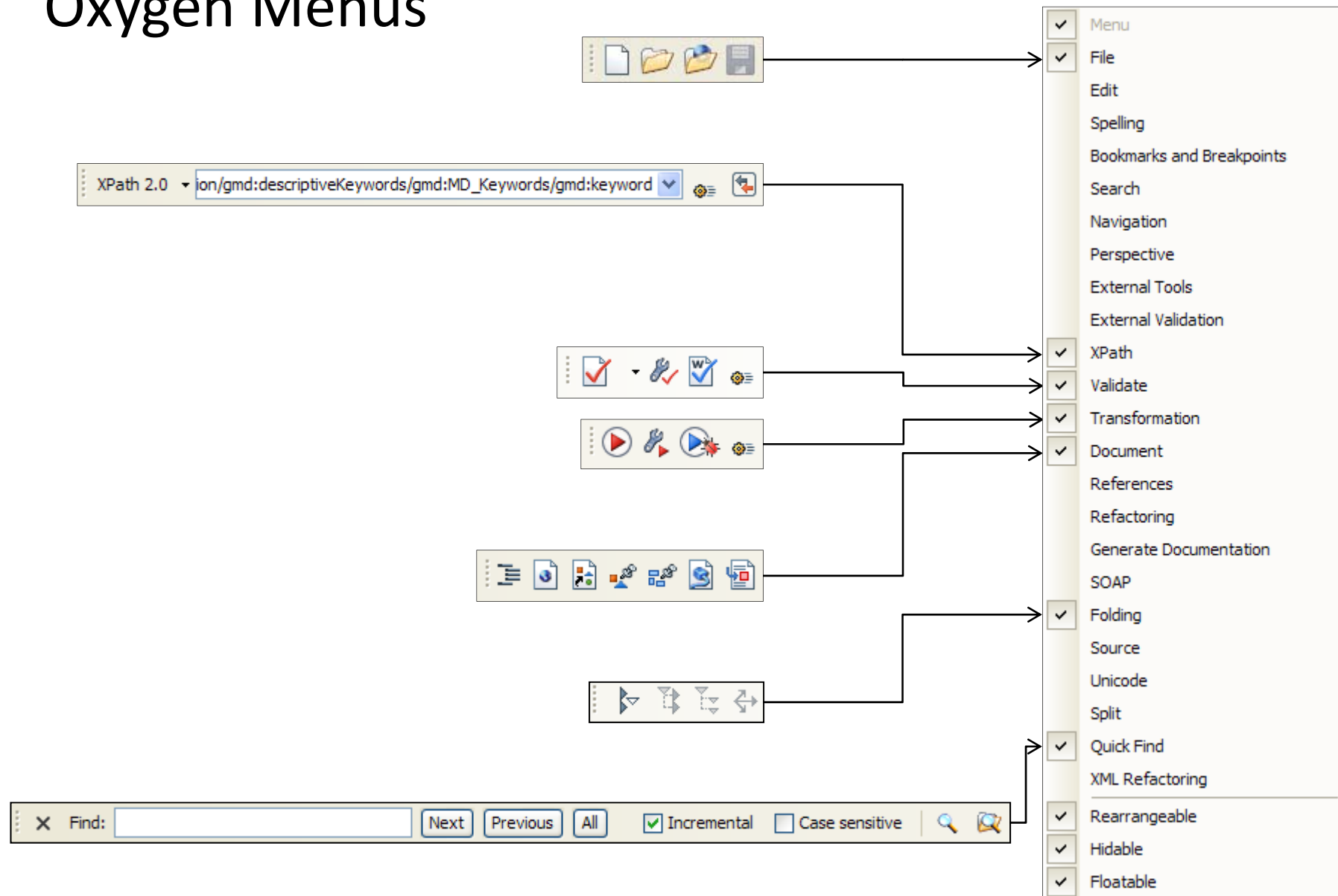
The bottom pane shows the 'Find' search bar with options for 'Next', 'Previous', 'All', 'Incremental', and 'Case sensitive'. The status bar at the bottom indicates 'Learn completed', 'U+003C', and '2:1'.

Organize Projects
Folders
Logical Folders
Batch Processing
Batch Search/Replace

Record Overview
Interactive Outline
XPath Search/Display

Edit
Auto-completion
Auto-validation
Auto-indent
Access to Definitions
Content Templates
CSS Interfaces

Oxygen Menus



CSS for any iso Metadata Record

The screenshot shows an XML editor window titled "1786.xml [C:\Documents and Settings\amilan\My Documents\1Metadata\1XML Stylesheets\FGDC to ISO - 2012\ISO resolved\1786.xml] - <oxyen/> XML Editor". The editor displays a metadata record for "1786.xml" with various fields highlighted in yellow. The fields are organized into sections: "File identifier:", "Language:", "Character set:", "Hierarchy level:", "Metadata author:", "Contact information:", "Role:", and "Date stamp:". The "File identifier" field contains "gov.noaa.ngdc.mgg.dem:1786". The "Language" field contains "eng: USA". The "Character set" field contains "MD_CharacterSetCode: utf8" and "@codeList: http://www.isotc211.org/2005/resources/CodeList/gmxCodeLists.xml#MD_CharacterSetCode, @codeListValue: utf8". The "Hierarchy level" field contains "MD_ScopeCode: dataset" and "@codeList: http://www.isotc211.org/2005/resources/CodeList/gmxCodeLists.xml#MD_ScopeCode, @codeListValue: dataset". The "Metadata author" section includes "CI_ResponsibleParty:" with "Individual name:" "Kelly Carignan" and "Organisation name:" "DOC/NOAA/NESDIS/NGDC gt National Geophysical Data Center, NESDIS, NOAA, U.S. Department of Commerce". The "Contact information" section includes "CI_Contact:" with "Address:" "Electronic mail address:" "kelly.carignan@noaa.gov". The "Role" section includes "CI_RoleCode: pointOfContact" and "@codeList: http://www.ngdc.noaa.gov/metadata/published/xsd/schema/resources/CodeList/gmxCodeLists.xml#CI_RoleCode, @codeListValue: pointOfContact". The "Date stamp" field is empty. The editor also shows a "Project" pane on the left and a "Text Grid Author" pane at the bottom.

1786.xml [C:\Documents and Settings\amilan\My Documents\1Metadata\1XML Stylesheets\FGDC to ISO - 2012\ISO resolved\1786.xml] - <oxyen/> XML Editor

File Edit Find Project Options Tools Author Document Window Help

xml-to-text-ISO.xsl* x 1786.xml x

gmi:MI_Metadata gmd:contact gmd:CI_ResponsibleParty gmd:individualName gco:CharacterString

MI_Metadata:

File identifier:

gov.noaa.ngdc.mgg.dem:1786

Language:

eng: USA

Character set:

MD_CharacterSetCode: utf8

(@codeList: http://www.isotc211.org/2005/resources/CodeList/gmxCodeLists.xml#MD_CharacterSetCode, @codeListValue: utf8)

Hierarchy level:

MD_ScopeCode: dataset

(@codeList: http://www.isotc211.org/2005/resources/CodeList/gmxCodeLists.xml#MD_ScopeCode, @codeListValue: dataset)

Metadata author:

CI_ResponsibleParty:

Individual name:

Kelly Carignan

Organisation name:

DOC/NOAA/NESDIS/NGDC gt National Geophysical Data Center, NESDIS, NOAA, U.S. Department of Commerce

Contact information:

CI_Contact:

Address:

CI_Address:

Electronic mail address:

kelly.carignan@noaa.gov

Role:

CI_RoleCode: pointOfContact

(@codeList: http://www.ngdc.noaa.gov/metadata/published/xsd/schema/resources/CodeList/gmxCodeLists.xml#CI_RoleCode, @codeListValue: pointOfContact)

(@uuid: dad4ff9a-31bd-47dc-a653-8ae3e7948944, @id:)

(@xlink:href: , @xlink:title: Kelly Carignan, @uuidref:)

Date stamp:

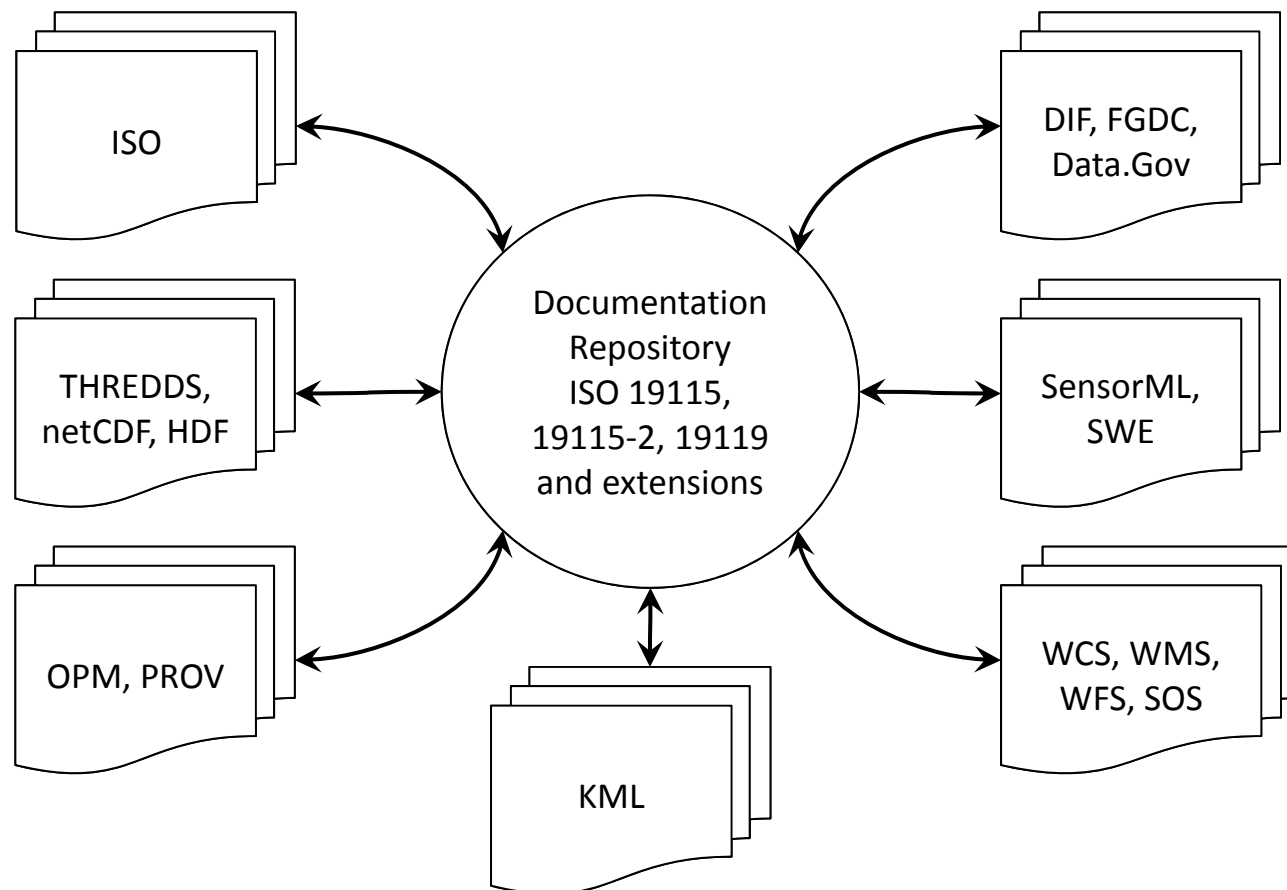
Text Grid Author

C:\...amilan\My Documents\1Metadata\1XML Stylesheets\FGDC to ISO - 2012\ISO resolved\1786.xml XPath - successful U+0000

start Mozilla Firefox "NGDC System No..." 1786.xml [C:\Doc... SecureFX C:\Documents an... C:\Documents an... Microsoft PowerP...

1:47 PM

Documentation in Multiple Dialects



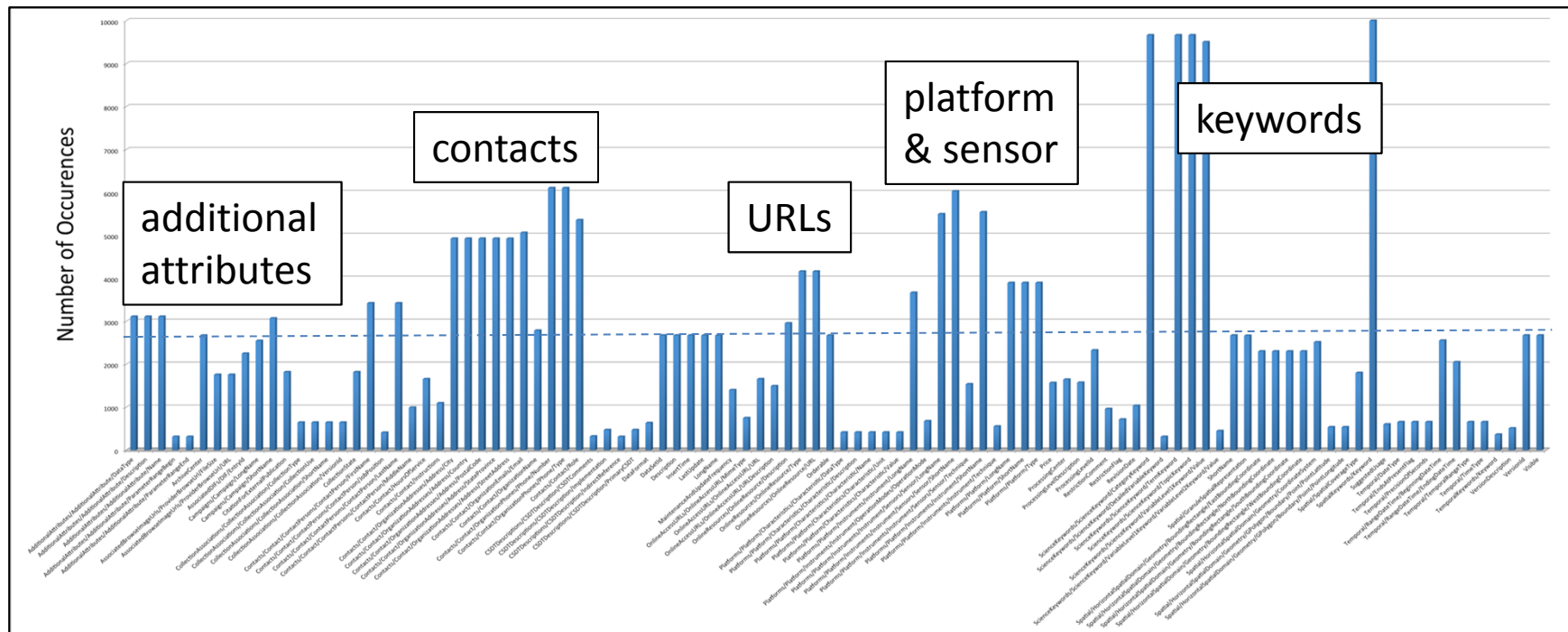
The global environmental community is currently using a variety of metadata dialects to document data and services. Translations between these dialects is, and will be, important for some time to come. This raises a number of questions...

Translation Proofing

Translations between dialects can be characterized by the amount of information that does or doesn't make it across the translation. This is the “lossyness” of the translation and it varies from 0 (nothing lost = good) to 1 (all lost = bad).

Translation Proofing is the process of determining the loss associated with translating a metadata **collection** from one dialect to another.

Why a collection?



ECHO Collections to ISO

Element	Count	%	Translate?	Translated
ArchiveCenter	2672	1.02%	Yes	2672
AssociatedBrowseImageUrls/ProviderBrowseUrl/FileSize	1758	0.67%	No	0
AssociatedBrowseImageUrls/ProviderBrowseUrl/URL	1758	0.67%	Yes	1758
AssociatedDIFs/DIF/EntryId	2250	0.86%	Yes	2250
Campaigns/Campaign/LongName	2548	0.97%	Yes	2548
Campaigns/Campaign/ShortName	3071	1.17%	Yes	3071
Campaigns/Campaign/StartDate	2	0.00%	No	0
Total	14059	100%		12299

Seven elements occur 14,059 times in 2673 records (all ECHO collection metadata)

Five translate, two don't: FileSize is more important than StartDate

71% of the elements and 87.5% of the content is translated

Complete sample: 262,422 content elements, 61% translate, 90.1% of content

NOAA ENC-Direct FGDC to ISO

111 elements occur 18,500 times in 158 records

90 translate, 21 don't

81% of the elements and 87% of the content is translated

Some non-standard content, some fields that don't translate (addrtype), some projection and coordinate system information is by reference in ISO

In the end 5% of the content will be left on the floor (addrtype, formcont)

Next?

Improve translation by including more elements (if possible)

Evaluate whether non-translated content is critical (StartDate?)

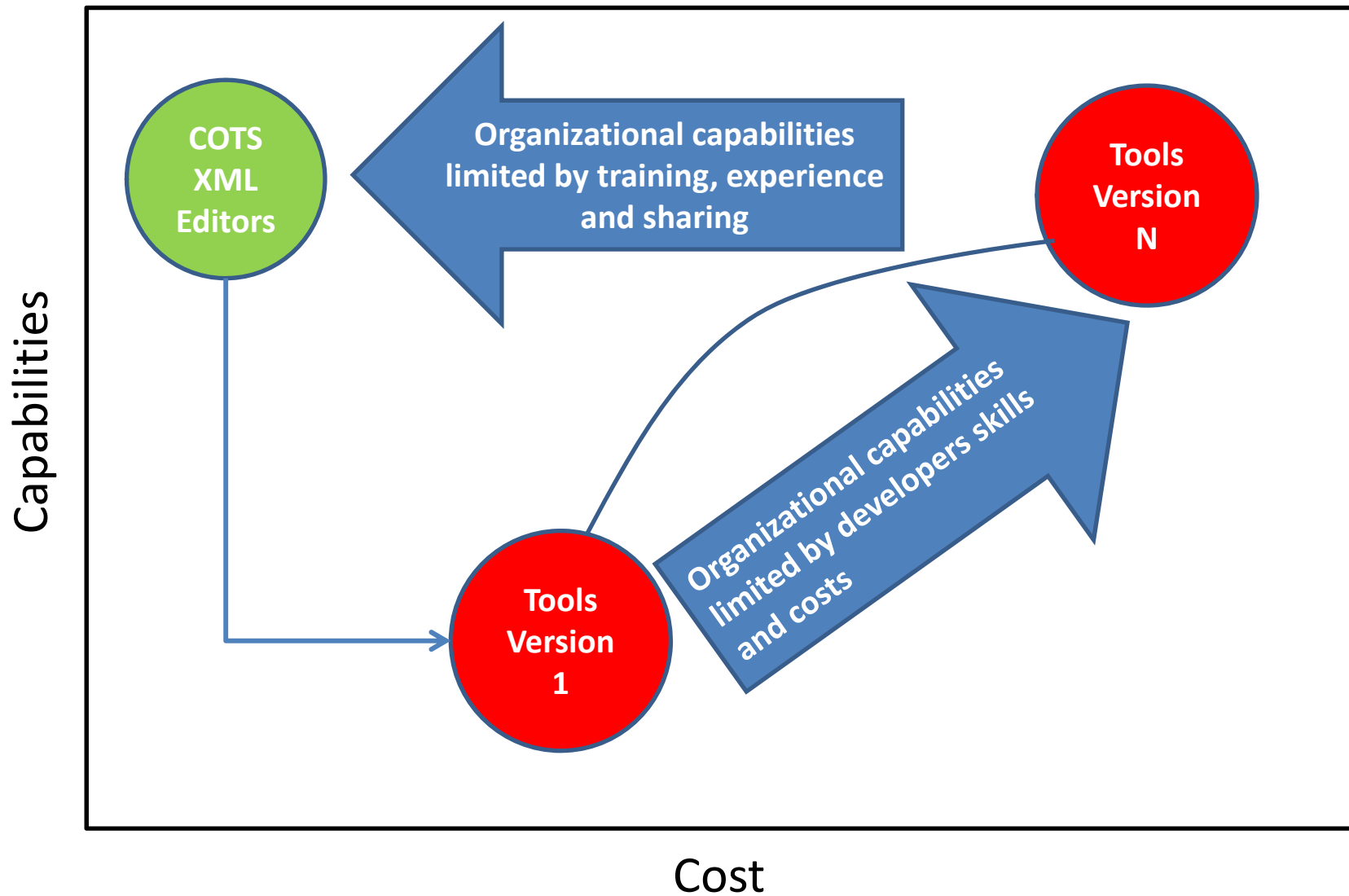
If no, is it critical in the source dialect?

Extend the target model to include more content

Revise target standard

Point back to original dialect and include translation version

Capabilities and Costs



Questions?

